British Mycological Society promoting fungal science

journal homepage: www.elsevier.com/locate/funbio

Analysis of microsatellite markers in the genome of the plant pathogen *Ceratocystis fimbriata*

*Melissa C. SIMPSON, P. Markus WILKEN, Martin P. A. COETZEE, Michael J. WINGFIELD, Brenda D. WINGFIELD**

Department of Genetics, Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, Private Bag X20, Hatfield, Pretoria 0028, South Africa

ARTICLE INFO

Article history:
Received 17 October 2012
Received in revised form
13 June 2013
Accepted 17 June 2013
Available online 26 June 2013
Corresponding Editor:
Michael Lorenz

Keywords:
Ceratocystis fimbriata
Genome mining
Microsatellites

ABSTRACT

Ceratocystis fimbriata sensu lato represents a complex of cryptic and commonly plant pathogenic species that are morphologically similar. Species in this complex have been described using morphological characteristics, intersterility tests and phylogenetics. Microsatellite markers have been useful to study the population structure and origin of some species in the complex. In this study we sequenced the genome of *C. fimbriata*. This provided an opportunity to mine the genome for microsatellites, to develop new microsatellite markers, and map previously developed markers onto the genome. Over 6000 microsatellites were identified in the genome and their abundance and distribution was determined. *Ceratocystis fimbriata* has a medium level of microsatellite density and slightly smaller genome when compared with other fungi for which similar microsatellite analyses have been performed. This is the first report of a microsatellite analysis conducted on a genome sequence of a fungal species in the order Microascales. Forty-seven microsatellite markers have been published for population genetic studies, of which 35 could be mapped onto the *C. fimbriata* genome sequence. We developed an additional ten microsatellite markers within putative genes to differentiate between species in the *C. fimbriata* s.l. complex. These markers were used to distinguish between 12 species in the complex.

Introduction

Microsatellites are 1–6 base pair tandem repeats that are abundant throughout eukaryotic and prokaryotic genomes (Tautz & Renz 1984; Field & Wills 1996). They are ideal molecular markers due to their high levels of polymorphism, they are inherited in a Mendelian manner and easily amplified using PCR (Levinson & Gutman 1987; Tautz 1989). Consequently, these markers have been used for strain typing, genetic mapping, and population structure studies in many different organisms (Field & Wills 1996; Jarne & Lagoda 1996; Hennequin *et al.* 2001). A major obstacle in using microsatellites for such studies is the need for their *de novo* isolation because hybridisation experiments using repeat-containing

* *Corresponding author*. Present address: Faculty of Natural and Agricultural Sciences, University of Pretoria, Private Bag X20, Hatfield, Pretoria 0028, South Africa. Tel.: +27 12 420 3946; fax: +27 12 420 3960.
E-mail addresses: Melissa.Simpson@fabi.up.ac.za (M. C. Simpson), Markus.Wilken@fabi.up.ac.za (P. M. Wilken), Martin.Coetzee@fabi.up.ac.za (M. P. A. Coetzee), Mike.Wingfield@fabi.up.ac.za (M. J. Wingfield), Brenda.Wingfield@fabi.up.ac.za (B. D. Wingfield).
1878-6146/$ – see front matter
http://dx.doi.org/10.1016/j.funbio.2013.06.004

probes can be inefficient and difficult to perform (Tautz & Renz 1984; Zane et al. 2002).

It has recently become possible to search sequence databases and whole genome sequences to estimate microsatellite distribution and abundance (Richard & Dujon 1996; Tóth et al. 2000; Demuth et al. 2007). This is a powerful tool for the development of informative markers for population studies (Drury et al. 2009). Knowledge of the position of microsatellites in the genome allows for unlinked microsatellites to be chosen (Selkoe & Toonen 2006). Such unlinked loci are essential for studies on populations in order to determine the variation within and between species as these loci are not associated with each other.

Genome-wide searches for microsatellites have been conducted for a number of eukaryotic organisms. These include *Arabidopsis thaliana*, *Drosophila melanogaster*, *Caenorhabditis elegans*, as well as the fungi *Saccharomyces cerevisiae*, *Neurospora crassa*, and *Fusarium graminearum* (Tóth et al. 2000; Katti et al. 2001; Lim et al. 2004; Karaoglu et al. 2005). The general consensus drawn from these studies is that fungal genomes contain fewer microsatellites than other higher eukaryotes. Furthermore, it is clear that every organism displays a unique microsatellite distribution with an abundance of certain types of microsatellite motifs (Tóth et al. 2000; Morgante et al. 2002). Since the initial research on microsatellites in fungi, there has been an increasing interest in the distribution and evolution of microsatellites in the genomes of these organisms (Tóth et al. 2000; Lim et al. 2004; Karaoglu et al. 2005).

In this study, we considered the microsatellite distribution in the fungal plant pathogen *Ceratocystis fimbriata*. This fungus was first described as the causal agent of sweet potato rot in 1890 (Halsted). Since then, many fungi have been identified as representing this species and infecting a wide variety of plants of agricultural and economic importance around the world, including coffee (Pontis 1951), poplar (Gremmen & de Kam 1977), *Acacia* species (Morris et al. 1993), and *Eucalyptus* species (Roux et al. 2004, 2000). Phylogenetic inference based on DNA sequence data has led to the recognition that *C. fimbriata sensu lato* represents a complex of cryptic species, some of which might be host-specific (Barnes et al. 2003; Engelbrecht & Harrington 2005; Johnson et al. 2005; van Wyk et al. 2009). Species in this complex have been named based on studies using DNA sequence comparisons, intersterility tests, and molecular markers (Engelbrecht & Harrington 2005; Johnson et al. 2005; Van Wyk et al. 2013).

Microsatellite markers have been useful for population studies of species in the *C. fimbriata* s.l. complex. They have provided insight into the population structure and origin of *Ceratocystis cacaofunesta*, *Ceratocystis platani*, and *Ceratocystis pirilliformis* (Barnes et al. 2003; Engelbrecht et al. 2007; Ocasio-Morales et al. 2007; Kamgan Nkuekam et al. 2009), which cause diseases on *Theobroma cacao*, *Platanus* species, and *Eucalyptus* trees, respectively. In some cases, microsatellite markers have also been used effectively to differentiate between isolates of *Ceratocystis* species from different geographical regions and hosts (Barnes et al. 2001). Most recently, 20 microsatellite markers were developed to differentiate between mango-associated isolates of *C. fimbriata* s.l. in Brazil (Rizatto et al. 2010).

Despite the application of microsatellites in a number of studies on Ascomycetes, little is known regarding the genome-wide abundance or distribution of microsatellites in *Ceratocystis* species, or even other species in the order *Microascales*, which accommodates *Ceratocystis*. The aim of this study was, therefore, to use a bioinformatics approach to determine the distribution and abundance of microsatellites in *C. fimbriata* and to compare the microsatellite structure of the *C. fimbriata* genome and that of other Ascomycete genomes. In addition, a set of microsatellite markers were developed and tested for their potential to recognise taxa in the *C. fimbriata* s. l. complex.

Methods and materials

Genome sequence and GC content of Ceratocystis fimbriata

DNA was extracted from a single isolate of *C. fimbriata* (CMW 14799/CBS114723, preserved in the culture collection of the Forestry and Agricultural Biotechnology Institute, University of Pretoria and the Centraalbureau voor Schimmelcultures, Utrecht) and subjected to genome sequencing. This was done using 454 pyrosequencing technology (Roche Diagnostics, Mannheim, Germany) at Inqaba Biotec (Pretoria, South Africa). The resulting reads were assembled into a draft genome using the Newbler version 2.3 genome assembler. To obtain information for the draft genome version, the 'create detailed mapping report' command of the CLC Genomics Workbench package version 5.0.1 (CLC Bio, Aarhus, Denmark) was used. In addition, the sequence statistics function was employed to produce a table of the nucleotide content of each contig. The GC content of the genome was determined from the CLC Genomics Workbench output files and calculated in Microsoft Office Excel 2007 (Microsoft Corp., Redmond, WA, USA).

Microsatellite discovery in the genome

Sequence files of the assembled *Ceratocystis fimbriata* genome were mined for microsatellite repeats using the online interface of MSatFinder version 2.0 (Thurston & Field 2005). A regex-directed search engine was used to identify sequences containing perfect microsatellites. This search engine is rapid and searches the sequence only once, but it cannot detect microsatellites of less than three repeats. A perfect microsatellite was defined for the search as a tract consisting of exact copies of the repeat unit, e.g. $(CTA)_6$, and one containing no mismatches or interruptions (Chambers & MacAvoy 2000). The minimum repeat number to detect mononucleotides was set to 12 repeats, while the minimum repeat number to detect di, tri, tetra, penta, and hexanucleotides was set to five. The microsatellites identified in MSatFinder were sorted in Microsoft Office Excel 2007 according to the type of motif and repeat length. The sequence lengths of each motif, the number of each type of motif, and the total repeats per megabase (Mb) of sequence analysed were calculated. In addition, the percentage of each type of microsatellite in the genome was analysed.

Microsatellites in coding regions

De novo prediction of open reading frames (ORFs) was performed using the online interface of AUGUSTUS (Keller *et al.* 2011). For this purpose, the annotated genome of *Fusarium graminearum* was used as the reference genome. The fasta output file containing sequences of the predicted coding genes was then searched for microsatellites and analysed using MSatFinder with the same settings as above.

Comparisons between fungal genomes

Microsatellite content in the *Ceratocystis fimbriata* genome was compared with that in the genomes of the Ascomycetes *Ashbya gossypii*, *Aspergillus fumigatus*, *Aspergillus nidulans*, *Candida albicans*, *Fusarium graminearum*, *Magnaporthe grisea*, and *Neurospora crassa*, as well as the Ascomycetous yeasts *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe* for which whole genome microsatellite analyses have been performed. Genome sizes and estimates of microsatellite abundance in each of the fungal genomes were obtained from the studies of Karaoglu *et al.* (2005) and Lim *et al.* (2004), and the GC contents were obtained from Lim *et al.* (2004).

Screening the Ceratocystis fimbriata *genome for published microsatellites*

Fasta files of the microsatellite sequences for the *C. fimbriata* s.l. species complex published by Barnes *et al.* (2001), Steimel *et al.* (2004), and Rizatto *et al.* (2010) were obtained from GenBank (Accession numbers in Supplementary Table 1). Using the CLC Genomics Workbench, local basic local alignment search tool (BLAST) searches of the microsatellite sequences against the *C. fimbriata* genome sequence were performed. Primer sequences for each microsatellite were obtained from the relevant studies (Barnes *et al.* 2001; Steimel *et al.* 2004; Rizatto *et al.* 2010). The binding sites of the primers were then identified using the primer function in CLC Genomics Workbench. Where a microsatellite could not be identified, a local BLAST search was performed on all the raw reads against the microsatellite sequences.

Contigs from the *C. fimbriata* genome containing the microsatellite sequences were analysed using the online interface of AUGUSTUS (Keller *et al.* 2011). The *Fusarium graminearum* genome sequence was used as the reference annotated genome to determine where the predicted genes were located. The microsatellite locations were then compared to the putative gene locations to determine their presence within coding regions, introns or noncoding regions. Proteins of the predicted genes that contain microsatellites were subsequently identified by analysing their amino acid sequences using BLASTp searches (Altschul *et al.* 1990).

Comparative analysis of microsatellites in Ceratocystis *spp*

Ninety isolates representing 26 species in the *Ceratocystis fimbriata* s.l. complex (Van Wyk *et al.* 2013) and one species (*Ceratocystis virescens*) residing in the *Ceratocystis coerulescens* complex (Wingfield *et al.* 2013; Supplementary Table 2) were grown on 2 % (v/w) malt extract agar (MEA, Biolab, Midrand, South Africa) supplemented with 100 mg L^{-1} streptomycin sulphate salt (SIGMA, Steinheim, Germany) and 100 mg L^{-1} thymine hydrochloride (SIGMA, Steinheim, Germany) for 2 weeks at 25 °C. Hyphal tips were isolated and grown on 2 % MEA. DNA extraction was performed as previously described by Van Wyk *et al.* (2006).

Microsatellite motifs, excluding mononucleotides, of five repeats or more were targeted to design the primers. Where a putative microsatellite was less than 50 bp from the either end of the contig, it was not analysed further. Using Primer3 (Rozen & Skaletsky 2000), primers were designed for 40 microsatellite loci. These primers were then tested on the genome sequence in CLC Genomics Workbench to determine whether they would result in the amplification of a single fragment. Primers for 36 microsatellite loci (Supplementary Table 3) were synthesized at Inqaba Biotec and then tested on four isolates representing different species in the *C. fimbriata* s.l. complex, including *Ceratocystis cacaofunesta* (CMW 26375), *Ceratocystis manginecans* (CMW 13851), *Ceratocystis platani* (CMW 1896), and *Ceratocystis polyconidia* (CMW 23818).

The PCR reactions were prepared using 20–50 ng of DNA in a 25 µl reaction containing one unit of MyTaq polymerase (Bioline Ltd., London, United Kingdom), 5× reaction buffer (consisting of 5 mM dNTPs and 15 mM $MgCl_2$), and 10 mM of each primer. These reactions were performed on an Eppendorf thermocycler (Eppendorf, Hamburg, Germany). The first denaturation step was carried out at 95 °C for 1 min, followed by 35 cycles of 95 °C for 15 s, 50 °C for 15 s, and 72 °C for 10 s, with a final elongation step of 72 °C for 7 min. The PCR products were then visualised on a 2 % (v/w) agarose gel under UV light.

The PCR products were purified with the DNA Clean and Concentrator™ Kit (Zymo Research Corporation, California, USA). Sequencing was performed in 10 µl reactions using either the forward or reverse primers for each microsatellite and a Big Dye cycle sequencing Kit v 3.1 (Perkin–Elmer, Warrington, UK) following the manufacturer's instructions. Sequencing PCR reactions were purified using the ZR DNA Sequencing Clean-Up™ Kit (Zymo Research Corporation, California, USA). An ABI PRISM 3300 Genetic Analyser was used to generate the sequences and the chromatograms that were produced were analysed using the computer programme BioEdit (Hall 1999). The sequences were aligned using MAFFT (Katoh & Toh 2008).

Ten microsatellite loci that showed polymorphisms between the four species were selected for screening using Genescan (Applied Biosystems, California, USA; Supplementary Table 3). The forward primer of each pair was resynthesized and fluorescently labelled with PET, 6-FAM, NED or VIC (Applied Biosystems, California, USA). PCR was then performed as above to amplify fragments in all isolates used in this study (Supplementary Table 2) but with annealing temperatures ranging from 42 °C to 50 °C and a final elongation step of 45 min. Four of the PCR products were then combined, each to a dilution of 1:100, according to their amplicon size and the type of fluorescent dye utilised. One microlitre of the mix was combined with 0.18 µL Genescan-600 Liz internal size standard (Applied Biosystems, California, USA) and 10 µL formamide. These mixes were then separated on a 36 cm capillary with POP™ polymer on an ABI PRISM 3100 Genetic Analyser. The Peak Scanner Software v1.0 (Applied

Biosystems, California, USA) was used to analyse the fragment sizes.

Results

Genome sequence and GC content of Ceratocystis fimbriata

Full genome sequencing of *C. fimbriata* isolate CMW 14799 resulted in ~2.5 million reads with an average read length of 248 bases, which was assembled into 3668 contigs (GenBank Accession number APWK01000000). The draft genome sequence that was produced had an estimated size of ~29.5 MB (as calculated by summation of all the contig sizes), a 20× average coverage, and an N50 contig size of 42 879 bases. The estimated GC content for the contigs of this fungus was 48.10 %. This GC content is almost equal to that of *Fusarium graminearum* (48.3 %) and is 2.2 % less than *Aspergillus nidulans*, 3.5 % less than *Magnaporthe grisea*, and 1.8 % less than *Neurospora crassa*. *Ceratocystis fimbriata* had a GC content that is higher than *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe* by 10.4 % and 12.1 %, respectively.

Abundance and characteristics of microsatellites

A total of 6737 perfect microsatellites were identified from the genome of *Ceratocystis fimbriata* using MSatFinder. The microsatellite density was estimated at one microsatellite for every 4.4 kilobase (Kb). The microsatellites were distributed among the six classes of microsatellites (Table 1). All types of mono, di, and trinucleotide motifs were present, but only some tetra, penta, and hexanucleotide motifs were found. For example, $(CAACAT)_n$ was not present in the genome. The trinucleotides $(CAT)_5$ and $(CAC)_5$ that have been used as DNA fingerprinting probes for a number of fungi, including species of *Ceratocystis* (DeScenzo & Harrington 1994; Harrington *et al.* 1998), were found in the *C. fimbriata* genome, four and eight times respectively.

AUGUSTUS predicted 8809 ORFs from the assembled contigs of *C. fimbriata*. From these, MSatFinder identified 739 microsatellites contained in the exons of the genes. The majority of microsatellites were trinucleotides, followed by dinucleotides and hexanucleotides (Table 2). Few mononucleotides were identified in the coding regions and these mainly consisted of C_n, which was found 15 times. The two most abundant trinucleotide motifs, $(CAG)_n$ and $(CAA)_n$, code for the amino acid glutamine while the third most abundant trinucleotide $(GCA)_n$ codes for alanine (Supplementary Table 4). The most abundant hexanucleotides $(CAGGCT)_n$, $(CAGGCA)_n$, and $(GCTCAA)_n$ code for glutamine/alanine tracts and $(CAGCAA)_n$ encodes glutamine. Other trinucleotide microsatellites that were fairly abundant in the genome encode for lysine, serine, and threonine.

Comparisons among fungal genomes

The genome size and GC content of the contig sequences of *Ceratocystis fimbriata* were similar to that in other Ascomycetes for which whole genome microsatellite analyses have been performed (Table 3). The most abundant microsatellites

Table 1 – The most abundant microsatellite motifsin the *C. fimbriata* genome, their density and total number.

Microsatellite class	Total motifs (% of all microsatellites)	Abundancy (microsatellites/Mb)	Abundant motifs	Number of motifs	% of each microsatellite class	% of total microsatellites
Mononucleotides	2995 (44.5%)	63	T	823	27.5	12.2
			A	770	25.7	11.4
			G	713	23.8	10.6
			C	689	23.0	10.2
Dinucleotides	2339 (34.7%)	49	AG/GA	565	24.1	8.39
			CT/TC	521	22.3	7.73
			AT/TA	474	20.3	7.03
			AC/CA	390	16.7	5.79
			GT/TG	362	15.5	5.37
			CG/GC	28	1.20	0.42
Trinucleotides	984 (14.6%)	21	ACG/AGC/CAG/CGA/GAC/GCA	167	17.0	2.48
			CGT/CTG/GCT/GTC/TCG/TGC	125	12.7	1.86
			AAG/AGA/GAA	93	9.45	1.38
			CTT/TCT/TTC	85	8.64	1.26
Tetranucleotides	208 (3.09%)	4	TGTA	9	4.33	0.13
			ATAC	8	3.85	0.12
			CCAG	6	2.88	0.09
			TACA	6	2.88	0.09
			AGAC	5	2.40	0.07
			GCTG	5	2.40	0.07
Pentanucleotides	110 (1.63%)	2	CAGCA	3	2.73	0.04
			TGTTT	3	2.73	0.04
Hexanucleotides	101 (1.50%)	2	TCTCTG	3	2.97	0.04
Total of all microsatellites	6737 (100%)	141	–	–	–	–

Table 2 – Total number of microsatellites and most abundant motifs in coding regions of the *C. fimbriata* genome.

Microsatellite class	Total motifs (% of all microsatellites)	Abundant motifs	Number of motifs	% of each microsatellite class	% of all microsatellites
Mononucleotides	16 (2.17%)	C	15	93.8	2.03
		G	1	6.30	0.14
Dinucleotides	154 (20.8%)	CT/TC	50	32.5	6.77
		AC/CA	43	27.9	5.82
		AG/GA	36	23.4	4.67
		CG/GC	14	9.09	1.89
		GT/TG	9	5.84	1.22
		AT/TA	2	1.30	0.27
Trinucleotides	511 (69.2%)	ACG/AGC/CAG/CGA/GAC/GCA	181	35.4	24.5
		AAC/ACA/CAA	89	17.4	12.0
		AAG/AGA/GAA	47	9.20	6.36
		CTG/GCT/TCG/TGC	46	9.00	6.22
Tetranucleotides	0	–	–	–	–
Pentanucleotides	1 (0.1%)	GACAG	1	100	0.13
Hexanucleotides	57 (7.7%)	CAGGCT	4	7.02	0.55
		CAGCAA	3	5.26	0.41
		CAGGCA	3	5.26	0.41
		GCTCAA	3	5.26	0.41
Total of all microsatellites	739 (100%)	–	–	–	–

among the Ascomycetes compared in this study were mononucleotides followed by di, tri, tetra, penta, and hexanucleotides (Table 4). Analysis of microsatellite lengths revealed that the longest microsatellites in *C. fimbriata* are, in most cases, shorter than those reported for some of the other Ascomycetes but are similar in length to *Aspergillus nidulans* (Table 5).

Mapping published microsatellites onto the Ceratocystis fimbriata *genome*

Thirty-four of the 47 published microsatellites developed for *Ceratocystis* species could be placed onto the *C. fimbriata* genome (Supplementary Table 1) and most of these were present on different contigs. Ten of these microsatellites were present in putative gene regions (Supplementary Table 5), however, some had a low coverage and/or a high *E*-value. The full sequences of four microsatellites (CfCAA80, CfCAG15, Cfim16, and Cfim18) could not be determined in the genome because they were located at the ends of the contigs.

The presence and locations of nine microsatellite sequences could not be determined in the *C. fimbriata* genome. Six microsatellites (Cfim01, Cfim02, Cfim05, Cfim08, Cfim13, and Cfim14) had similar motifs, i.e. the motif $(AC)_{14}$ was present in three of the microsatellites and the rest consisted of various $(AC)_n$ or $(TG)_n$ repeats, as well as similar flanking regions. These six similar microsatellite sequences all aligned to the same position in the *C. fimbriata* genome. The other three microsatellites (Cfim03, Cfim06, and Cfim07), with $(AC)_n$ motifs, also aligned to the same position although their flanking regions were not as similar as the other microsatellites. Analyses of the raw reads failed to resolve these microsatellites further because there was little or no coverage of the flanking regions, and the raw reads that did cover parts of the microsatellite sequence had low quality scores.

Table 3 – Comparison of the genome size, GC content, microsatellite density, and total microsatellites in the genomes of various Ascomycetes.

Organism	Sequence analysed (Mb)	GC content	Microsatellite density	Total microsatellites
A. gossypii[b]	8.7	52.0[b]	1/0.34 kb	25 526
A. fumigatus[b]	29.4	49.6[b]	1/0.36 kb	82 173
A. nidulans[a]	30.1	50.3[b]	1/12.5 kb	2410
C. albicans[b]	14.9	33.5[b]	1/0.12 kb	129 076
F. graminearum[a]	36.1	48.3[b]	1/12.5 kb	2896
M. grisea[a]	37.9	51.6[b]	1/3.3 kb	11 642
N. crassa[a]	38.0	49.9[b]	1/2.7 kb	14 319
S. cerevisiae[a]	14.2	37.7[b]	1/3.9 kb	3618
Sch. pombe[a]	13.1	36.0[b]	1/4.4 kb	3232
C. fimbriata	29.5	48.1	1/4.4 kb	6737

a Results from Karaoglu *et al.* (2005).
b Results from Lim *et al.* (2004).

Table 4 – Comparison of the distribution of microsatellites in the genomes of various Ascomycetes.

Organism	Mono (%)	Di (%)	Tri (%)	Tetra (%)	Penta (%)	Hexa (%)
A. gossypii[b]	97.8	1.40	0.75	0.01	0.01	0.03
A. fumigatus[b]	97.9	1.35	0.66	0.05	0.02	0.04
A. nidulans[a]	51.8	31.2	13.5	1.49	1.16	0.79
C. albicans[b]	95.4	1.62	2.48	0.28	0.08	0.09
F. graminearum[a]	38.3	35.7	19.2	2.94	2.52	1.38
M. grisea[a]	69.1	14.9	13.5	1.88	0.28	0.39
N. crassa[a]	41.5	22.4	28.5	5.29	1.34	0.94
S. cerevisiae[a]	65.3	22.6	11.0	0.39	0.25	0.53
Sch. pombe[a]	72.2	20.5	6.06	0.65	0.56	0.09
C. fimbriata	44.5	34.7	14.6	3.09	1.63	1.50

a Results from Karaoglu et al. (2005)
b Results from Lim et al. (2004).

Microsatellites CF17/18 and CF23/24, with $(CA)_{15}$ and $(TG)_{15}$ motifs respectively, aligned well to the same position in the *C. fimbriata* genome. Both their primers had 100 % similarity within this region. These two microsatellite sequences appeared to be the same sequence as they aligned well to one another once one sequence was reverse complemented. Another two microsatellites, Cfim09 and Cfim17, had the same motif $(AC)_9$, and aligned well to one another except for the end flanking regions of the sequences. They both aligned to the same position in the *C. fimbriata* genome. However, the alignment of Cfim17 to the contig was better than that of Cfim09. Only the reverse primers had some similarity to this region. Locations of all these microsatellites could not be determined further, even after analysing the raw reads.

Comparative analysis of microsatellites in Ceratocystis *spp*

Thirty-six microsatellites within gene regions (Supplementary Table 6) were chosen as markers. Coding regions are expected to share some similarity between species, such that the markers could be used on more than one species (Metzgar *et al.* 2000). Most of the putative microsatellites chosen were trinucleotides, with some dinucleotides, tetranucleotides, hexanucleotides, and one octanucleotide.

After amplification of the 36 microsatellites, one primer pair (CF_GTT6) was discarded because it produced two fragments in all the isolates tested. The remainder of the microsatellite primers produced single amplicons. Nine microsatellites showed no polymorphisms in any of the isolates tested. The rest had at least two alleles with six showing different alleles in four species, including *Ceratocystis fimbriata*. Most of the loci present on the same contigs had different allele complements between the four different species tested but some loci were monomorphic. Single nucleotide polymorphisms (SNPs) were identified at nine loci, some of which disrupted the microsatellite motifs, but in essence retained the same repeat number. From the 26 polymorphic microsatellites, ten loci were chosen for further analysis using Genescan; one contained an SNP (CF_CAGAAG5) but still showed differences in the microsatellite repeat number.

The ten polymorphic primers were tested on isolates representing different species in the *C. fimbriata s.l.* species complex and an outgroup species, *Ceratocystis virescens* (Table 6). Some primer pairs failed to amplify loci in all the isolates representing a species and these were considered to be null alleles. The ten polymorphic markers produced a total of 141 alleles with a size range of 123 bp–360 bp. The smallest number of alleles per locus was eight (CF_CAAG5) and the largest was 20 (CF_GCT11). The locus with the lowest allele diversity is found in the microsatellite designed within an intron of a putative gene.

Discussion

Microsatellite analysis

Microsatellites have been characterised in various Ascomycete fungi (Lim *et al.* 2004; Karaoglu *et al.* 2005; Li *et al.* 2009) but this is the first study in which these repeat units have been characterised across the genome of a member of the *Microascales* and more specifically in *Ceratocystis fimbriata*. *Ceratocystis fimbriata* was chosen for this study because it is

Table 5 – Comparison of the longest motifs in each different group of microsatellite motifs in the genomes of various Ascomycetes.

Organism[a]	Mono	Di	Tri	Tetra	Penta	Hexa
A. gossypii[b]	$(A)_{57}$	–	$(AAC)_{21}$	–	$(AAAAT)_{23}$	–
A. fumigatus[b]	$(A)_{95}$	$(AG)_{26}$	$(AAG)_{28}$	$(AAAG)_{28}$	–	$(AACCCT)_{28}$
A. nidulans[a]	$(T)_{94}$	$(GT)_{36}$	$(TGA)_{31}$	$(AAAT)_{13}$	$(AAACG)_{14}$	$(TTAGGG)_{22}$
C. albicans[b]	$(A)_{35}$	–	$(AAT)_{30}$	$(AAAT)_{29}$	–	$(AGAGCC)_{20}$
F. graminearum[a]	$(T)_{41}$	$(CT)_{28}$	$(GAA)_{46}$	$(CTTT)_{13}$	$(GTATG)_{18}$	$(TGAAGA)_{22}$
M. grisea[a]	$(T)_{59}$	$(GA)_{92}$	$(TGG)_{37}$	$(TACC)_{48}$	$(GGCAA)_{29}$	$(GCCTGA)_{58}$
N. crassa[a]	$(T)_{89}$	$(TC)_{78}$	$(TTA)_{93}$	$(AGGA)_{51}$	$(AAGGA)_{32}$	$(AGGGTT)_{28}$
S. cerevisiae[a]	$(T)_{42}$	$(GA)_{32}$	$(TAT)_{36}$	$(AAAT)_{13}$	$(GATGA)_{7}$	$(TGTTTT)_{8}$
Sch. pombe[a]	$(T)_{39}$	$(TG)_{19}$	$(CAA)_{28}$	$(TAAA)_{7}$	$(TATTT)_{9}$	$(ATTATC)_{6}$
C. fimbriata	$(G)_{62}$	$(GA)_{41}$	$(AAG)_{19}$	$(TCAC)_{15}$	$(GACAG)_{18}$	$(GAAAAT)_{14}$

– indicates that no data was available.
a Results from Karaoglu *et al.* (2005).
b Results from Lim *et al.* (2004).

Table 6 – Genescan analysis of microsatellite loci fragment sizes for each isolate of *Ceratocystis* used in this study. Grey boxes indicate conflicting allele sizes within a species, x indicates no PCR product, and 0 indicates double bands.

Species	CMW	CF_CAA/CAG80	CF_CAAG5	CF_GTT50	CF_GAT5	CF_CAG AAG5	CF_GCT 11	CF_CTCTC TGT5	CF_CAA/CAG81	CF_CCG5	CF_CAA/CAG24
C. acaciivora	22562	x	x	x	236.7	231.8	x	x	221.4	x	173.9
C. acaciivora	22563	227.9	196.7	x	236.8	x	x	x	221.5	x	173.5
C. acaciivora	22595	227.9	189	x	236.9	x	x	236.9	221.4	190.2	173.7
C. acaciivora	22621	227.9	x	233.5	236.8	220	x	237.1	221.5	190.3	167.7
C. albifundus	4068	x	x	254.3	248.3	196.4	x	x	218.8	189.1	156.1
C. albifundus	4090	x	x	254.1	x	196.8	x	x	218.7	189.2	152.5
C. albifundus	5329	x	x	245.2	x	231.7	237.6	x	218.7	189.6	153
C. albifundus	14159	x	x	x	x	196	x	190.3	x	189.3	x
C. albifundus	15760	x	192.9	248.3	254	207.4	238	x	215.6	189.3	152.6
C. atrox	19385	246	x	248.4	x	208.7	201.4	153.3	218.4	181.1	122.8
C. atrox	19389	242.5	x	247.9	x	208.6	189.1	194.5	218.3	181.2	123.2
C. cacaofunesta	14809	222.1	205.1	230.1	236.8	196.8	237.4	194.3	224.3	190.6	159
C. cacaofunesta	15051	222	197.4	230.1	228.4	196.5	237.7	194.1	224.3	190	158.9
C. cacaofunesta	26375	222.1	195.6	233.1	232.7	208.2	238.2	194	230.2	196.3	161.9
C. caryae	14793	x	x	223.9	x	200	348.3	x	250.4	x	156.5
C. caryae	14808	x	x	223.7	245.9	x	304.8	x	256.4	x	x
C. colombiana	5751	239.5	196.7	236.8	239.8	244.5	234.9	267.9	232.9	x	182.8
C. colombiana	5761	239.7	197.1	236.5	240	244.4	234.9	267.8	233	187	182.7
C. colombiana	9565	248.6	197.6	236.5	239.8	231.6	x	275.9	232.9	187.4	182.7
C. colombiana	11280	239.6	197.6	236.9	239.8	226	234.4	307.9	232.8	187.4	182.8
C. curvata	22433	225.1	192.6	233.1	246.1	214.4	256.2	240.7	221.8	190.2	167.7
C. curvata	22435	225.1	193.1	233.2	246.1	213.4	256	241.1	221.5	190.4	168
C. curvata	22442	225	193.4	232.9	246	213.8	256.3	241	226.3	190.3	168
C. diversiconidia	22445	222.2	184.7	242.1	236.9	238.7	219.2	180	215.2	189.8	155.7
C. diversiconidia	22446	222.3	185	242.1	236.8	239.1	219.8	179.6	215.4	190.8	156
C. diversiconidia	22448	222.2	184.9	242	236.9	239.1	219.4	179.9	215.1	190.3	155.7
C. ecuadoriana	22092	239.8	192.9	236.8	248.3	206	244.3	185.9	218.4	190.2	164.5
C. ecuadoriana	22093	230.8	193.2	235.9	249.3	205.4	244.2	186	218.3	190	164.4
C. ecuadoriana	22097	230.8	193.6	236.2	249.3	205.9	244	185.9	218.3	190.3	165
C. ecuadoriana	22405	230.9	192.8	236.5	240	205.8	244.4	185.9	218.3	190	165.1
C. fimbriata	1547	225	193.2	233.3	239.9	214	231.1	234.7	224.3	187.5	168
C. fimbriata	14799	225	193.4	233.2	240.6	214.2	231.6	234.6	224.4	187.4	167.9
C. fimbriata	15049	225.1	193	232.9	239.6	214.4	231.6	234.6	x	187.1	167.6
C. fimbriatomima	24174	221.9	200.7	233.3	239.9	208.5	237.2	x	209.6	190	170.9
C. fimbriatomima	24176	222	200.7	233.7	x	208.3	237.3	192.3	224.3	190.2	170.5
C. fimbriatomima	24377	222	200.9	233.3	240	208.3	237.3	192.4	224.3	190.1	170.9
C. fimbriatomima	24378	222.1	200.8	232.9	240	208.5	237	192.3	224.2	190.3	170.6
C. fimbriatomima	24379	221.7	200.9	233.5	x	x	237.3	x	224.3	190.5	171.2
C. larium	25434	242.5	191.8	233.3	233.8	208.5	188.9	194.6	217.7	203.5	x
C. larium	25435	242.5	x	233.5	234.7	208.7	189.1	x	217.5	203.7	149.6
C. larium	25436	242.5	192.9	233.2	234.4	208.8	188.9	x	x	203.6	150.3
C. larium	25437	242.5	192.8	233.3	234.5	x	188.9	194.5	218.7	203.6	149.7
C. manginecans	13851	227.9	197.1	233.8	236.8	220.2	240.4	236.9	221.3	190.6	174
C. manginecans	13852	227.7	x	233.7	236.3	219.9	240.4	236.9	221.4	190.3	173.6
C. manginecans	15314	228	197.1	233.2	236.5	220.3	240.6	236.9	221.3	190.2	173.8
C. manginecans	15317	228	196.6	233.6	236.5	220.2	240.4	236.8	221.4	190.2	173.5
C. manginecans	23634	227.9	197	233.2	236.6	220.2	240.5	236.9	221.3	190.1	173.4
C. neglecta	11285	x	185.2	239.3	236.4	203.5	246.9	208.3	227.1	190.5	161.8
C. neglecta	11284	230.8	185.2	239.3	236.7	219.9	247.2	202.6	227.1	190.4	161.9
C. neglecta	18194	227.8	192.9	236.3	x	214.5	243.8	186.4	215.5	190	165
C. neglecta	17808	230.8	193.2	236.1	248.2	214.3	243.8	186	215.6	190.2	162.1
C. obpyriformis	23806	227.9	193.2	286.9	228.4	196.8	232.2	153.3	212.5	180.9	122.3
C. obpyriformis	23807	x	x	x	228.1	196.3	x	x	212.4	180.9	230

(continued on next page)

Table 6 (*continued*)

Species	CMW	CF_CAA/CAG80	CF_CAAG5	CF_GTT50	CF_GAT5	CF_CAGAAG5	CF_GCT11	CF_CTCTCTGT5	CF_CAA/CAG81	CF_CCG5	CF_CAA/CAG24
C. papillata	8850	x	x	x	227.9	x	237.4	x	x	x	x
C. papillata	8856	230.9	196.9	235.9	239.5	x	222.4	179.9	218.4	x	171.1
C. papillata	8857	x	x	x	228	237.7	x	x	x	x	x
C. papillata	28662	239.4	197.1	235.8	239.6	x	x	x	x	199.2	170.2
C. pirilliformis	6579	x	x	248.3	228.4	202.5	x	153.4	248.3	181.1	122.8
C. pirilliformis	6583	233.9	x	280.1	228.1	202.5	232.4	153	209.6	181.4	122.3
C. pirilliformis	12671	233.8	x	274.6	231.2	202.7	x	153.4	209.6	180.9	122.5
C. pirilliformis	16511	233.9	197.2	274.6	231.4	202.8	232	153.5	251.3	181.1	122.9
C. pirilliformis	28200	x	185.4	280.3	x	x	233	x	218.7	x	173.7
C. platani	1896	230.7	185.7	238.9	236.8	219.8	247.6	202.5	227.2	190	162.1
C. platani	14802	233.7	185.4	239.7	236.6	220.4	247	202.6	227.2	190.1	161.6
C. platani	23450	230.8	285.3	239.6	236.6	219.9	246.9	202.3	227.2	189.9	161.8
C. platani	23918	230.8	185.7	239.4	236.3	220.2	247.1	x	227.2	190.3	161.5
C. platani	26380	233.7	184.9	239.8	236.7	x	247.3	202.5	227.1	190	161.8
C. polychroma	11424	227.7	x	251.4	x	x	201.5	152.2	215.3	181.9	167.8
C. polychroma	14281	x	x	251.3	x	199.1	x	x	218.2	181.9	167.5
C. polyconidia	23818	225	196.9	233.5	236.8	214	271	226.7	221.5	189.9	x
C. populicola	14789	224.8	x	223.2	x	257.3	256.6	x	223.6	207.3	132.8
C. smalleyi	14800	242.9	x	224.1	227.9	x	360.3	x	216.7	181	141.2
C. tanganyicensis	15992	222.3	192.8	x	231.2	178.6	183.2	156.1	218.8	193.5	152.6
C. tanganyicensis	15999	221.9	192.1	x	x	178.3	x	156.2	x	x	152.9
C. tsitsikammensis	13982	224.8	185	245.1	246	208.9	232	171.8	218.8	187.3	173.8
C. tsitsikammensis	14275	224.9	185.2	245.3	245.9	208.9	192.5	171.4	230.7	187.4	173.6
C. tsitsikammensis	14276	224.8	185	245.4	245.6	209	192.5	171.4	218.6	187.4	173.2
C. tsitsikammensis	14280	225	185.4	245.3	246	208.8	192.3	171.5	218.4	187.3	173.6
C. variospora	20935	222.3	146.4	208.2	268.9	218.4	384.5	x	227.2	211.5	161.9
C. variospora	26384	x	x	x	228.2	241.4	231	x	x	x	x
C. variospora	26386	222.3	x	x	268.6	x	x	x	228.7	211.4	164.7
C. virescens	3225	x	x	233.3	228.2	x	0	0	x	x	x
C. virescens	11160	x	196.9	x	x	185.3	0	0	x	x	x
C. virescens	11164	x	197	233.2	228.3	x	0	0	x	219.6	x
C. virescens	17335	x	x	x	228.3	x	0	0	x	219.8	x
C. virescens	17339	x	x	233.1	x	x	0	0	x	219.6	x
C. zombamontana	15235	233.9	192.8	251.4	231.5	196.6	231.5	153.2	209.7	181.4	122.3
C. zombamontana	15236	233.8	197.3	250.9	231.1	196.6	x	x	253.8	190	123

the type species of the genus *Ceratocystis* and it is one of a complex of important plant pathogens of forestry and agricultural crops worldwide (Kile 1993; Roux & Wingfield 2009). Results of the analyses are consistent with those of previous studies that have shown that microsatellites are less abundant in fungi and that genome size does not correlate with microsatellite density (Tóth *et al.* 2000; Lim *et al.* 2004; Karaoglu *et al.* 2005).

Generally, fungi contain fewer and shorter microsatellites than other eukaryotes, and in this regard *C. fimbriata* was no exception (Tóth *et al.* 2000; Lim *et al.* 2004; Karaoglu *et al.* 2005; Li *et al.* 2009). The genome of this fungus is almost the same size as *Aspergillus fumigatus* (29.4 Mb) but larger than that of the yeasts *Saccharomyces cerevisiae* (14.2 Mb) and *Schizosaccharomyces pombe* (12.1 Mb) for which whole genome microsatellite analyses have been done (Lim *et al.* 2004; Karaoglu *et al.* 2005). The lengths of the longest microsatellites in *C. fimbriata* are mostly shorter than the longest microsatellites in other Ascomycetes (Lim *et al.* 2004; Karaoglu *et al.* 2005). This suggests that the microsatellites in *C. fimbriata* could be imperfect and passing through the death phase (i.e. the microsatellite sequence is decaying). The resulting microsatellites are therefore, separated into shorter perfect tracts and unique sequences through the accumulation of mutations (Ellegren 2000). Alternatively, the shorter microsatellite lengths could be the result of evolutionary constraints, such as genome size or the sequences adjacent to the microsatellite that prevent the formation of long tracts of microsatellites.

We found that the *C. fimbriata* genome had a medium microsatellite density compared to that in other fungi (Tóth *et al.* 2000; Lim *et al.* 2004; Karaoglu *et al.* 2005). It has been suggested that differences in genome organisation and efficiency of the mismatch repair machinery could contribute to variation in microsatellite density among fungi (Tóth *et al.* 2000; Karaoglu *et al.* 2005). In addition to these factors, GC content can also affect microsatellite density (Lim *et al.* 2004). Lim *et al.* (2004) showed that fungal genomes with more abundant microsatellites have a lower GC content, whereas those with a 50 % GC content had a more equal distribution of microsatellites. This is clearly seen with the yeasts, *S. cerevisiae* and *Sch. pombe*, which have a low GC content and a high microsatellite density (Lim *et al.* 2004; Karaoglu *et al.* 2005). Results of the present study revealed approximately 50 % GC content for *C. fimbriata* with a medium microsatellite density, which is not without precedence, as this has also been observed in *Neurospora crassa* (Lim *et al.* 2004).

Ceratocystis fimbriata and *A. fumigatus* have similar genome sizes and GC content, however, their microsatellite densities differ greatly, with the former species having a medium density and the latter a very high density of microsatellites. The density of microsatellites in *Ashbya gossypii* and *Candida albicans* is also very high (Lim *et al.* 2004). This can be attributed to the different software and constraints used to identify microsatellites (Merkel & Gemmell 2008). Lim *et al.* (2004) searched for microsatellites with at least five repeats for all classes while Karaoglu *et al.* (2005) identified microsatellites in all classes that were 10 bp or more. In the present study, mononucleotides of 12 or more repeats and at least five repeats of di–hexanucleotides were identified. The results of

Lim *et al.* (2004) are biased towards mononucleotides and thus reflect a much higher overall density, making it difficult to provide a detailed comparison between studies.

Several studies, including this one, have shown that patterns of microsatellite distribution are similar among fungi (Harr *et al.* 2002; Lim *et al.* 2004; Karaoglu *et al.* 2005). However, there are differences with regards to the most abundant motifs within each microsatellite class that makes each organism unique (Tóth *et al.* 2000; Lim *et al.* 2004; Karaoglu *et al.* 2005). In the *C. fimbriata* genome, trinucleotides represent the third most abundant class of microsatellites overall and the most abundant class within coding regions. Trinucleotides, along with hexanucleotides, encode amino acids such as glutamine and asparagine that are commonly found in fungi and other eukaryotes (Tóth *et al.* 2000; Li *et al.* 2009). The abundance of these particular triplet repeats in coding regions is thought to be linked to selection acting on the amino acid repeats in proteins (Albà *et al.* 1999).

Tri and hexanucleotides are expected to be the most abundant microsatellites in coding regions because they would not change the reading frame and consequently the amino acid sequence of the resulting protein (Metzgar *et al.* 2000). Similarly, other motifs with total lengths being multiples of three are also expected to be present (Metzgar *et al.* 2000; Gibbons & Rokas 2009). It was, therefore, surprising that the longest mononucleotide (C_{14}) was found four times and had a length of 14 bases, which is not a multiple of three. This particular motif is likely tolerated in the predicted coding regions as it could have particular properties required by the resulting proteins. However, because this is only an *in silico* predicted coding region, it is possible that it does not form part of a gene. Most of the microsatellites (81 %) in coding regions had lengths that are a multiple of three, which correlates well with the fact that the trinucleotides are the most abundant microsatellite class in coding regions However, the second most abundant class, dinucleotides, found in coding regions did not occur in copy numbers totalling a length that is a multiple of three. An increase or decrease in the number of repeats for these microsatellites may not be tolerated because changes in the resulting protein could adversely affect the growth and development of the fungus.

Mapping the published microsatellites onto the Ceratocystis fimbriata *genome*

The *C. fimbriata* genome did not contain all of the microsatellites published for *Ceratocystis* species. The fact that these microsatellite regions were not identified in *C. fimbriata* possibly reflects the variability of these markers between species. The ten published microsatellites (Barnes *et al.* 2001; Steimel *et al.* 2004; Rizatto *et al.* 2010), which have now been identified as being within putative gene regions, are probably not ideal for studies on populations or genetic mapping. This is because neutral markers are required for such studies and these are thus normally present in noncoding regions (Selkoe & Toonen 2006).

Primers developed for microsatellites in noncoding regions are not likely to be transferable to related species because these regions are not under selection and are thus more variable (Barbará *et al.* 2007). This has been seen in population studies where some of the microsatellites developed for one species are monomorphic or the primers do not function in another species within the *C. fimbriata* s.l. complex (Barnes *et al.* 2005; Engelbrecht *et al.* 2007, 2004; Ocasio-Morales *et al.* 2007; Kamgan Nkuekam *et al.* 2009; Ferreira *et al.* 2010). For example, microsatellite primers developed for *C. fimbriata* were not all transferable to *Ceratocystis albifundus* and *Ceratocystis pirilliformis* (Barnes *et al.* 2005; Kamgan Nkuekam *et al.* 2009).

Microsatellites CF17/18 and CF23/24 shared the same motif (although in the reverse complement), flanking regions and aligned to the same position in the *C. fimbriata* genome. Using these two markers together is, therefore, not recommended because they would result in the same data. The other microsatellite markers that are similar to each other were all developed by Rizatto *et al.* (2010). These 13 microsatellite markers are problematic as they have similar, if not the same, motifs (AC, CA or TG) and flanking regions, and align to the same position in the *C. fimbriata* genome. Results of the analyses in the present study and the fact that other microsatellites are available, lead us to conclude that the latter markers should not be used for all species in the *C. fimbriata* s.l. complex.

Comparative analysis of microsatellites in Ceratocystis *spp*

The 25 polymorphic microsatellites identified in this study consisted mostly of trinucleotides but also included dinucleotides, hexanucleotides, a tetranucleotide, and an octanucleotide. One dinucleotide and the tetranucleotide were present within introns while the rest were found in coding regions. The tri and hexanucleotides showed the most polymorphisms, probably because they are triplet repeats and the loss or gain of a repeat unit would not disrupt the reading frame (Metzgar *et al.* 2000). Also, constraints might not be placed on the number of repeated amino acids in a homopolymer tract because the protein function may not necessarily be influenced by a change in repeat number.

The ten microsatellite markers developed in this study were used to differentiate between a number of species in the *Ceratocystis fimbriata* s. l. complex. These markers can thus be used for interspecies comparisons. Those species in the complex that are phylogenetically most closely related, showed more consistent allele sizes for all the loci. Some species showed different patterns of amplification and allele sizes at some loci, which might indicate diversity within the species and could thus be used for interspecies comparisons (Supplementary Table 7). However, additional data would be required to use these markers with confidence due to the fact that only one or two isolates were tested for some species and mixed results were obtained for others. It is likely that there have been mutations at the microsatellite loci whereby the microsatellites themselves or primer binding sites have changed, or even where sections of the microsatellite regions have been removed. This could occur, not only between species (which would explain the fact that microsatellite markers are not always transferable across species) but also within species.

Markers other than microsatellites, for example SNPs, also show polymorphisms and can differentiate between species. Some of the microsatellites tested in this study showed SNPs

not only in the flanking regions but also in the microsatellite motifs themselves. SNPs present in the microsatellite motifs disrupt the microsatellite but the overall length of the microsatellite is maintained and would be indistinguishable if only fragment sizes are analysed. These SNPs provide additional tools to differentiate between species in the *C. fimbriata* s.l. complex if they are found to be present in all isolates of a single species. Further analysis will be needed to verify that the SNPs identified are present within more than one isolate of each species.

Conclusions

The microsatellite content in the 28 Mb genome of *Ceratocystis fimbriata* is typical for eukaryotes. However, the distribution and abundance of microsatellites was unique among the fungal genomes that have been studied previously (Lim *et al.* 2004; Karaoglu *et al.* 2005). Generally, *C. fimbriata* has shorter microsatellites than other Ascomycetes, which could be attributed to its smaller genome size. The smaller genome size along with a medium microsatellite density was not surprising because previous studies have shown that genome size and microsatellite density are not always correlated (Lim *et al.* 2004; Karaoglu *et al.* 2005).

More than 6000 microsatellites were identified in *C. fimbriata* and these were found in both the coding and noncoding regions. Microsatellite markers have already been successfully used for studies on populations of species in the *C. fimbriata* complex (Barnes *et al.* 2001; Engelbrecht *et al.* 2007). The ten microsatellite markers developed in this study could differentiate between 12 species in the *C. fimbriata* s.l. complex. However, additional research is required in order to differentiate between all the species in the complex with confidence. These ten microsatellite markers, along with the many identified from the *C. fimbriata* genome, could be used in future studies for more robust population genetic studies and for species differentiation.

Acknowledgements

We thank the National Research Foundation (NRF), the South African Biosystematics Initiative (SABI), members of the Tree Protection Co-Operative Programme (TPCP), and the Department of Science and Technology (DST)/ NRF Centre of Excellence in Tree Health Biotechnology (CTHB), South Africa for funding this project.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.funbio.2013.06.004.

REFERENCES

Albà MM, Santibanez-Koref MF, Hancock JM, 1999. Amino acid reiterations in yeast are overrepresented in particular classes of proteins and show evidence of a slippage-like mutational process. *Journal of Molecular Evolution* **49**: 789–797.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ, 1990. Basic local alignment search tool. *Journal of Molecular Biology* **215**: 403–410.

Barbará T, Palma-Silva C, Paggi GM, Bered F, Fay MF, Lexer C, 2007. Cross-species transfer of nuclear microsatellite markers: potential and limitations. *Molecular Ecology* **16**: 3759–3767.

Barnes I, Gaur A, Burgess T, Roux J, Wingfield BD, Wingfield MJ, 2001. Microsatellite markers reflect intra-specific relationships between isolates of the vascular wilt pathogen *Ceratocystis fimbriata*. *Molecular Plant Pathology* **2**: 319–325.

Barnes I, Nakabonge G, Roux J, Wingfield BD, Wingfield MJ, 2005. Comparison of populations of the wilt pathogen *Ceratocystis albifundus* in South Africa and Uganda. *Plant Pathology* **54**: 189–195.

Barnes I, Roux J, Wingfield BD, Dudzinski MJ, Old KM, Wingfield MJ, 2003. *Ceratocystis pirilliformis*, a new species from *Eucalyptus nitens* in Australia. *Mycologia* **95**: 865–871.

Chambers GK, MacAvoy ES, 2000. Microsatellites: consensus and controversy. *Comparative Biochemistry and Physiology* **126**: 455–476.

Demuth JP, Drury DW, Peters ML, van Dyken D, Priest NK, Wade MJ, 2007. Genome-wide survey of *Tribolium castaneum* microsatellites and description of 509 polymorphic markers. *Molecular Ecology Notes* **7**: 1189–1195.

DeScenzo RA, Harrington TC, 1994. Use of $(CAT)_5$ as a DNA fingerprinting probe for fungi. *Phytopathology* **84**: 534–540.

Drury DW, Siniard AL, Wade MJ, 2009. Genetic differentiation among wild populations of *Tribolium castaneum* estimated using microsatellite markers. *Journal of Heredity* **100**: 732–741.

Ellegren H, 2000. Microsatellite mutations in the germline: implications for evolutionary inference. *Trends in Genetics* **16**: 551–558.

Engelbrecht CJB, Harrington TC, 2005. Intersterility, morphology and taxonomy of *Ceratocystis fimbriata* on sweet potato, cacao and sycamore. *Mycologia* **97**: 57–69.

Engelbrecht CJB, Harrington TC, Alfenas AC, Suarez C, 2007. Genetic variations in populations of the cacao wilt pathogen, *Ceratocystis cacaofunesta*. *Plant Pathology* **56**: 923–933.

Engelbrecht CJB, Harrington TC, Steimel J, Capretti P, 2004. Genetic variation in eastern North American and putatively introduced populations of *Ceratocystis fimbriata f. platani*. *Molecular Ecology* **13**: 2995–3005.

Ferreira EM, Harrington TC, Thorpe DJ, Alfenas AC, 2010. Genetic diversity and interfertility among highly differentiated populations of *Ceratocystis fimbriata* in Brazil. *Plant Pathology* **59**: 721–735.

Field D, Wills C, 1996. Long, polymorphic microsatellites in simple organisms. *Proceedings of the Royal Society of London: Biological Sciences* **263**: 209–215.

Gibbons JG, Rokas A, 2009. Comparative and functional characterisation of intragenic tandem repeats in 10 *Aspergillus* genomes. *Molecular Biology and Evolution* **26**: 591–602.

Gremmen J, de Kam M, 1977. *Ceratocystis fimbriata*, a fungus associated with poplar canker in Poland. *European Journal of Forest Pathology* **7**: 44–47.

Hall TA, 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acid Symposium Series* **41**: 95–98.

Halsted BD, 1890. Some fungous diseases of the sweet potato. The black rot. *Bulletin of the New Jersey Agricultural College Experiment Station* **76**: 7–14.

Harr B, Todorova J, Schlötterer C, 2002. Mismatch repair-driven mutational bias in *D. melanogaster*. *Molecular Cell* **10**: 199–205.

Harrington TC, Steimel J, Kile G, 1998. Genetic variation in three *Ceratocystis* species with outcrossing, selfing and asexual reproductive strategies. *European Journal of Forest Pathology* **28**: 217–226.

Hennequin C, Thierry A, Richard GF, Lecointre G, Nguyen HV, Gaillardin C, Dujon B, 2001. Microsatellite typing as a new tool

for identification of *Saccharomyces cerevisiae* strains. *Journal of Clinical Microbiology* **39**: 551–559.
Jarne P, Lagoda PJL, 1996. Microsatellites, from molecules to populations and back. *Trends in Ecology and Evolution* **11**: 424–429.
Johnson JA, Harrington TC, Engelbrecht CJB, 2005. Phylogeny and taxonomy of the North American clade of the *Ceratocystis fimbriata* complex. *Mycologia* **97**: 1067–1092.
Kamgan Nkuekam G, Barnes I, Wingfield MJ, Roux J, 2009. Distribution and population diversity of *Ceratocystis pirilliformis* in South Africa. *Mycologia* **101**: 17–25.
Karaoglu H, Lee CMY, Meyer W, 2005. Survey of simple sequence repeats in completed fungal genomes. *Molecular Biology and Evolution* **22**: 639–649.
Katoh K, Toh H, 2008. Recent developments in the MAFFT multiple sequence alignment program. *Briefings in Bioinformatics* **9**: 286–298.
Katti MV, Ranjekar PK, Gupta VS, 2001. Differential distribution of simple sequence repeats in eukaryotic genome sequences. *Molecular Biology and Evolution* **18**: 1161–1167.
Keller O, Kollmar M, Stanke M, Waak S, 2011. A novel hybrid gene prediction method employing protein multiple sequence alignments. *Bioinformatics* **27**: 757–763.
Kile GA, 1993. Plant diseases caused by species of *Ceratocystis sensu stricto* and *Chalara*. In: Wingfield MJ, Seifert KA, Webber JA (eds), *Ceratocystis and Ophiostoma: taxonomy, ecology and pathogenicity*. American Phytopathological Society, St. Paul, Minnesota, pp. 173–183.
Levinson G, Gutman GA, 1987. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Molecular Biology and Evolution* **4**: 203–221.
Li C, Liu L, Yang J, Li J, Su Y, Zhang Y, Wang Y, Zhu Y, 2009. Genome-wide analysis of microsatellite sequence in seven filamentous fungi. *Interdisciplinary Sciences and Computer Life Sciences* **1**: 141–150.
Lim S, Notley-McRobb L, Lim M, Carter DA, 2004. A comparison of the nature and abundance of microsatellites in 14 fungal genomes. *Fungal Genetics and Biology* **41**: 1025–1036.
Merkel A, Gemmell NJ, 2008. Detecting microsatellites in genome data: variance in definitions and bioinformatic approaches cause systematic bias. *Evolutionary Bioinformatics* **4**: 1–6.
Metzgar D, Bytof J, Wills C, 2000. Selection against frameshift mutations limits microsatellite expansion in coding DNA. *Genome Research* **10**: 72–80.
Morgante M, Hanafey M, Powell W, 2002. Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. *Nature Genetics* **30**: 194–200.
Morris MJ, Wingfield MJ, de Beer C, 1993. Gummosis and wilt of *Acacia mearnsii* in South Africa caused by *Ceratocystis fimbriata*. *Plant Pathology* **42**: 814–817.
Ocasio-Morales RG, Tsopelas P, Harrington TC, 2007. Origin of *Ceratocystis platani* on native *Platanus orientalis* in Greece and its impact on natural forests. *Plant Disease* **91**: 901–904.
Pontis R, 1951. A canker disease of the coffee tree in Colombia and Venezuela. *Phytopathology* **41**: 179–184.
Richard GF, Dujon B, 1996. Distribution and variability of trinucleotide repeats in the genome of the yeast *Saccharomyces cerevisiae*. *Gene* **174**: 165–174.
Rizatto S, de Araùjo Batista CE, Bajay MM, Sigrist MS, Ito MF, Monteiro M, Cavallari MM, Pinheiro JB, Zucchi MI, 2010. A new set of microsatellite markers for the genetic characterisation of *Ceratocystis fimbriata*, an economically important plant pathogen. *Conservation Genetics Resources* **2**: 55–58.
Roux J, van Wyk M, Hatting H, Wingfield MJ, 2004. *Ceratocystis* species infecting stem wounds on *Eucalyptus grandis* in South Africa. *Plant Pathology* **53**: 414–421.
Roux J, Wingfield MJ, 2009. *Ceratocystis* species: emerging pathogens of non-native plantation *Eucalyptus* and *Acacia* species. *Southern Forests* **71**: 115–120.
Roux J, Wingfield MJ, Bouillet JP, Wingfield BD, Alfenas AC, 2000. A serious new wilt disease of Eucalyptus caused by *Ceratocystis fimbriata* in Central Africa. *Forest Pathology* **30**: 175–184.
Rozen S, Skaletsky HJ, 2000. Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S (eds), *Bioinformatic Methods and Protocols: methods in molecular biology*. Humana Press, Totowa, New Jersey, pp. 365–386.
Selkoe KA, Toonen RJ, 2006. Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. *Ecology Letters* **9**: 615–629.
Steimel J, Engelbrecht CJB, Harrington TC, 2004. Development and characterisation of microsatellite markers for the fungus *Ceratocystis fimbriata*. *Molecular Ecology Notes* **4**: 215–218.
Tautz D, 1989. Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic Acids Research* **17**: 6463–6471.
Tautz D, Renz M, 1984. Simple sequences are ubiquitous repetitive components of eukaryotic genomes. *Nucleic Acids Research* **12**: 4127–4138.
Thurston MI, Field D, 2005. *MsatFinder: detection and characterisation of microsatellites*. CEH, Oxford, Mansfield Road, Oxford OX1 3SR, Distributed by the authors at. http://www.genomics.ceh.ac.uk/msatfinder/.
Tóth G, Gáspári Z, Jurka J, 2000. Microsatellites in different eukaryotic genomes: survey and analysis. *Genome Research* **10**: 967–981.
Van Wyk M, van der Merwe NA, Roux J, Wingfield BD, Kamgan GN, Wingfield MJ, 2006. Population genetic analyses suggest that the *Eucalyptus* fungal pathogen *Ceratocystis fimbriata* has been introduced into South Africa. *South African Journal of Science* **102**: 259–263.
Van Wyk M, Wingfield BD, Wingfield MJ, 2013. *Ceratocystis* species in the *Ceratocystis fimbriata* species complex. In: Seifert K, de Beer ZW, Wingfield MJ (eds), *The Ophiostomatoid Fungi: expanding frontiers*. *Centraalbureau voor Schimmelcultures. Biodiversity Series*, vol. 90, pp. 65–76.
van Wyk M, Wingfield BD, Clegg PA, Wingfield MJ, 2009. *Ceratocystis larium* sp nov., a new species from *Styrax benzoin* wounds associated with incense harvesting in Indonesia. *Persoonia* **22**: 75–82.
Wingfield BD, van Wyk M, Roos H, Wingfield MJ, 2013. Ceratocystis: emerging evidence for discrete generic boundaries. In: Seifert K, de Beer ZW, Wingfield MJ (eds), *The Ophiostomatoid Fungi: expanding frontiers*. *Centraalbureau voor Schimmelcultures*. *Biodiversity Series*, vol. 90, pp. 57–64.
Zane L, Bargelloni L, Patarnello T, 2002. Strategies for microsatellite isolation: a review. *Molecular Ecology* **11**: 1–16.